Journal of Advanced Biomedical Sciences
https://jabs.fums.ac.ir/
Online ISSN: 2783-1523
Fasa University of Medical Sciences

An Efficient Model based on the k-Nearest Neighbors Algorithm for Parkinson's Disease Detection

Mohammadjavad Hosseinpoor[1,2]

1. Department of Computer Engineering, Estahban Branch, Islamic Azad University, Estahban, Iran
2. Young Researchers and Elite Club, Estahban Branch, Islamic Azad University, Estahban, Iran

Article Info

Article Type:
Original Article

Article history:
Received
13 Nov 2024
Received in revised form
23 Feb 2025
Accepted
02 Mar 2025
Published online
10 Mar 2025

Publisher
Fasa University of Medical Sciences

Abstract

Background & Objectives: Parkinson's disease (PD) is a neurological disorder characterized by the progressive loss of brain cells, significantly affecting body movement. Early diagnosis not only reduces healthcare costs but also helps prevent adverse outcomes for patients. Researchers are increasingly utilizing intelligent machine learning methods to enhance the accuracy and efficiency of PD diagnosis.

Materials & Methods: Although several data mining techniques have achieved reasonable accuracy in diagnosing PD, they often encounter trade-offs between accuracy and execution speed and are sensitive to parameter settings and data outliers. The k-Nearest Neighbors (KNN) algorithm, for example, is valued for its simplicity and speed but suffers from limitations such as sensitivity to neighborhood size and reliance on majority voting, both of which can degrade performance. To address these challenges, this study employs an advanced variant of the KNN algorithm, referred to as Multiple Local Mean Vector-based Nearest Neighbor Classification (MLMV-NNC), alongside a neural network classifier trained using Bayesian backpropagation. The MLMV-NNC method enhances traditional KNN by incorporating multiple local mean vectors, thereby reducing the influence of outliers and improving classification robustness.

Results: The proposed diagnostic approach demonstrates superior performance in detecting PD. Specifically, the model achieves an accuracy of 99%, precision of 96%, specificity of 98.6%, and sensitivity of 100%. Furthermore, a comparative analysis with traditional methods, including Support Vector Machines (SVM) and Artificial Neural Networks (ANN), highlights the superior performance of the proposed method.

Conclusion: The findings indicate that the combination of MLMV-NNC and a neural network trained via Bayesian backpropagation constitutes a highly effective approach for diagnosing PD. This method not only improves accuracy but also mitigates common challenges such as sensitivity to parameter settings and data outliers, offering a promising alternative to conventional classification techniques.

Keywords: Parkinson's disease, Diagnosis, Data Mining, ANN, KNN, Accuracy, SVM

Cite this article: Hosseinpoor MJ. An Efficient Model based on the k-Nearest Neighbors Algorithm for Parkinson's Disease Detection. J Adv Biomed Sci. 2025; 15(2): 148-157.
DOI: 10.18502/jabs.v15i2.18083

Introduction

Parkinson's Disease (PD) is a progressive neurological disorder that affects approximately 1% of the population over the age of 55, with its prevalence increasing significantly with age (1). The disease is characterized by the degeneration of dopaminergic neurons in the substantia nigra pars compacta (SNc) and the presence of Lewy bodies, which are intracellular inclusions primarily composed of α-synuclein and ubiquitin (2).

Corresponding Author: Mohammadjavad Hosseinpoor, Department of Computer Engineering, Estahban Branch, Islamic Azad University, Estahban, Iran.
Email: mj.hosseinpoor@iau.ac.ir

This neuronal loss leads to a depletion of dopamine in the striatum, which disrupts both motor and non-motor functions. Clinically, PD manifests with symptoms such as tremors, bradykinesia (slowness of movement), rigidity, and postural instability, as well as non-motor symptoms including sleep disturbances, speech variability, and loss of smell (3). While the exact mechanisms underlying PD remain incompletely understood, oxidative stress, mitochondrial dysfunction, and neuroinflammation are widely recognized as key contributors to neurodegeneration (4). Early diagnosis is critical, as PD is irreversible and progressively worsens, often leading to severe disability and a reduced quality of life (5).

In recent years, machine learning (ML) and artificial intelligence (AI) techniques have emerged as promising tools for enhancing PD diagnosis (6). Traditional diagnostic methods, which rely on clinical observations and subjective assessments, can be enhanced by integrating ML algorithms capable of analyzing complex datasets. For instance, gait analysis, a non-invasive and cost-effective approach, has shown potential in detecting PD through changes in walking patterns (7). Similarly, ML models can process diverse data types, such as speech recordings, MRI images, and wearable sensor data, to detect early signs of the disease with higher accuracy (8). However, existing methods face significant challenges, including high computational costs, dependency on large datasets, and limitations in managing multidimensional data structures (9).

Techniques such as regression, k-Nearest Neighbor (k-NN), and Artificial Neural Networks (ANNs) often struggle with issues such as early convergence, local minima (10), and sensitivity to parameter settings (11). Recent advancements in AI have introduced more sophisticated approaches, such as Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs), which offer higher accuracy in PD diagnosis (12). For example, Abumalloh et al. (13) demonstrated that deep learning (DL) methods outperform traditional ML approaches in identifying PD, particularly when analyzing large datasets. Their bibliometric analysis highlighted the growing global interest in DL-based diagnostic methods, though they identified a research gap related to incremental learning techniques for big data analysis. Similarly, Sigcha et al. (14) reviewed the use of wearable devices combined with DL algorithms for continuous monitoring of PD symptoms. Their study, which analyzed 69 research papers, found that inertial sensors and CNNs were the most commonly used tools, with motor symptoms being more frequently studied than non-motor symptoms. Despite the potential of these technologies, challenges such as data variability and the need for standardized validation methods remain.

Tanveer et al. (15) conducted a comprehensive review of DNNs for PD diagnosis, noting their high accuracy but also their computational expense and hardware requirements. They emphasized the need for cost-effective solutions to facilitate the widespread clinical adoption of these technologies. Vyas et al. (16) proposed a 3D CNN model for analyzing brain MRI images, achieving an accuracy of 88.9%, which outperformed their 2D CNN model. Their approach utilized advanced preprocessing techniques, such as bias field correction and Z-score normalization, to enhance feature extraction from MRI data. Rajalaxmi et al. (17) introduced an improved binary grey wolf optimizer (BIGWO) for feature selection in PD diagnosis, demonstrating enhanced classification performance through adaptive k-NN (AkNN). Their work highlighted the importance of optimizing feature selection to boost diagnostic accuracy.

Despite these advancements, challenges such as computational inefficiency, data dependency, and the need for costly hardware persist. To address these limitations, this study proposes a novel hybrid approach that combines the

[DOI: 10.18502/jabs.v15i2.18083]

strengths of Multiple Local Mean Vector-based k-Nearest Neighbor (MLM-KHNN) and neural networks. MLM-KHNN improves upon traditional k-NN by addressing issues associated with feature weighting and majority voting, while neural networks enhance feature processing and relationship mapping. This hybrid model aims to optimize accuracy, precision, and computational efficiency, offering a standalone, offline diagnostic tool for early PD detection. By integrating the advantages of both methods, this approach seeks to overcome the limitations of existing techniques and provide a more reliable, cost-effective solution for clinical applications.

Methods and Materials

In this research, we examine a method for diagnosing PD using classification techniques such as ANN and the Multiple Local Mean (MLM)-KHNN algorithm. To apply data mining in diagnostic areas like disease detection, machine learning algorithms must be utilized to process information stored in databases. Various machine learning methods exist in this domain, and in this study, due to the criteria of speed, simplicity, and quick execution time, we focused on approaches that are both efficient and computationally feasible.

One of the most prominent, simple, and fast machine learning methods in data mining is KNN and the ANN approach (15). Despite its simplicity and speed, the KNN method faces significant challenges, such as dependence on majority voting, the difficulty of selecting an appropriate value for the neighborhood size parameter (k) (which requires extensive trial and error), and the equal treatment of all data points, regardless of their proximity to the query point. To address these limitations, in 2017, Zebin Pan et al. (2) introduced an advanced version of this algorithm, which, through multi-neighborhood selection and averaging multiple vectors, mitigates these challenges. This version not only retains simplicity and speed but also demonstrates enhanced performance in appropriately handling data with both linear and non-linear features.

Consequently, this research integrates the advanced version of KNN with a multilayer perceptron neural network (ANN) equipped with a Bayesian backpropagation algorithm for Parkinson's disease diagnosis. Machine learning methods, as intelligent data-driven approaches, require the division of data into two subsets: training and testing. In the training phase, both data features and decision-making features (class labels) are considered, while in the testing phase, only the data features are included. Classification methods are trained using the full information of the training data, including both data features and decision-making features, so that the classifier can evolve into a well-informed model. The testing data are then processed by the classifiers, and predictions are made regarding class labels (indicating the presence or absence of the disease). The predicted labels are compared with the actual labels to calculate classification errors and overall accuracy.

According to Figure 1, the detailed steps of the classification method for diagnosing Parkinson's disease are described as follows:

Data: he Parkinson's disease database, extracted from the UCI repository, was used. The dataset was split into training and testing sets in a 70:30 ratio to initiate the proposed diagnostic method.

Training the Classifier: The ensemble classifier based on ANN and MLM-KHNN was trained using the training data, enabling the classifier to develop into a well-informed model capable of diagnosing the disease when presented with data features. The ANN and MLM-KHNN classifiers were combined to optimize the diagnosis of Parkinson's disease. The relevant data were fed into the proposed combined method for the learning process. The MLM-KHNN method (multi-local mean nearest neighbors) was used for classification.

[DOI: 10.18502/jabs.v15i2.18083]

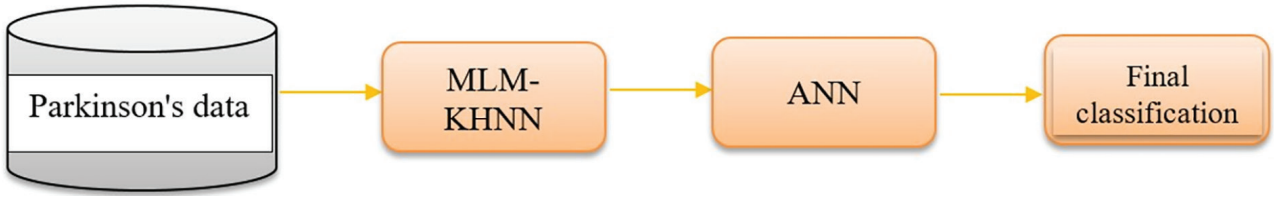

Figure 1. The Group Classification Process

To optimize the learning process, the output of the MLM-KHNN method was fed into a neural network with 20 hidden layers. While this number of layers facilitates deeper data processing, careful tuning of neural network parameters was performed to minimize overfitting and maximize computational efficiency.

Output Section: The final classification of the data was performed based on two categories: healthy individuals and Parkinson's patients.

The detailed workings of each classifier (MLM-KHNN and ANN) for Parkinson's disease diagnosis are explained as follows. Each of these classification methods was used to classify high-dimensional feature vectors due to their greater accuracy and precision in handling such data. The primary concept of the MLM-KHNN classifier is derived from the classification of classes by training feature vectors from various database sets. The extracted feature vectors were used as input to the classifier, and the data were randomly divided into two parts: training and testing. The training set was used to train the classifier, while the testing set was used to validate its performance. Feature computation and classification using MLM-KHNN were conducted in MATLAB. The k-fold cross-validation method was used to divide the dataset into training and testing sections. There are different methods for splitting the dataset, such as k-fold, 70:30, or 90:10 splits. In this study, the k-fold method was used for the MLM-KHNN algorithm. Evaluation is a tool for assessing the research method by dividing the dataset into training and testing sections, where the entire dataset is split into (k) parts. Each time, one of the (k) parts (a fold) is considered the test set, while the remaining (k-1) parts are used as the training set. In cross-validation, the data samples are first divided into (k) parts, ensuring that the volume of these parts is approximately equal whenever possible. Among these (k) parts, one is set aside as the test set, while the remaining (k-1) parts are used as the training set. Using this model, predictions are made on the test set data, and the model's accuracy is calculated using an appropriate loss function. Then, another of the (k) parts is chosen as the test set, and the process is repeated. Consequently, every data sample in the original set is predicted once.

Dataset

In this research, the Parkinson's disease database available on the UCI website was used to test the proposed data mining method based on the MLM-KHNN+ANN ensemble classifier for the diagnosis of Parkinson's disease. The database contains 22 data features (independent variables) and one output (dependent variable) named "Status," with two values: "disease" (1) and "healthy" (0). The details of the features and output of the database are provided in Table 1. This dataset is derived from a wide range of biomedical acoustic measurements taken from 195 individuals, including 147 patients with Parkinson's disease (PD) and 48 healthy individuals. Each column in the table represents a specific acoustic measurement, while each row corresponds to one of the 197 recorded sounds from these individuals. The data labels are assigned as 1 for PD patients and 0 for healthy individuals (1–5, 13–15).

Results

Simulation Environment and Parameters

The proposed method was implemented using

[DOI: 10.18502/jabs.v15i2.18083]

Table 1. Parkinson's disease Database from UCI

Number	Variable name	Parameter
1	MDVP: Fo(Hz)	Average audio frequency
2	MDVP: Fhi(Hz)	Maximum audio frequency
3	MDVP: Flo(Hz)	Audio frequency minimum
4	MDVP: Jitter(%)	Audio frequency changes in percentage
5	MDVP: Jitter(Abs)	Audio frequency changes in microsecond
6	MDVP: RAP	Audio frequency changes of relative amplitude
7	MDVP: PPQ	Audio frequency changes
8	Jitter:DDP	Audio frequency changes
9	MDVP: shimmer	Features domain changes
10	MDVP: shimmer(Db)	
11	Shimmer: APQ3	
12	Shimmer: APQ5	
13	Shimmer: APQ	
14	Shimmer: DDA	
15	NHR	Quality of presence of noise
16	HNR	Noise quality
17	DFA	Detrended DFA volatility analysis
18	RPDE	Entropy density measurement
19	Spread1	Nonlinear measurement of frequency
20	Spread2	
21	D2	Correlation rate
22	PPE	Non-linearity of PPE frequency
23	Status	A value of 0 indicates a healthy person and a value of 1 indicates a person with Parkinson's disease

Abbreviations: MDVP: Multi-Dimensional Voice Program, Fo(Hz): Fundamental Frequency, Fhi(Hz): Maximum Frequency, Flo(Hz): Minimum Frequency, Jitter(%): Jitter (percent), Jitter(Abs): Jitter (absolute), RAP: Relative Amplitude Perturbation, PPQ: Pitch Period Perturbation, DDP: Differential Deviation of Pitch, Shimmer: Shimmer, Shimmer(Db): Shimmer (decibels), APQ3: Amplitude Perturbation Quotient (3rd order), APQ5: Amplitude Perturbation Quotient (5th order), APQ: Amplitude Perturbation Quotient, DDA: Differential Deviation of Amplitude, NHR: Noise-to-Harmonics Ratio, HNR: Harmonics-to-Noise Ratio, DFA: Detrended Fluctuation Analysis, RPDE: Recurrence Period Density Entropy, Spread1: Nonlinear measure of frequency variation, Spread2: Nonlinear measure of frequency variation, D2: Correlation Dimension, PPE: Pitch Period Entropy, Status**:** A value of 0 indicates a healthy person and a value of 1 indicates a person with Parkinson's disease.

MATLAB R2024b and executed on a system with 16 GB **of** RAM, a 7-core processor, and Windows 10 (64-bit). The ANN classification algorithm was developed using MATLAB's Neural Network Toolbox, with the relevant neural network codes generated and customized according to the study's requirements. The key parameter settings for the implemented classification methods are summarized in Table 2.

Table 2 outlines the essential configurations used for training the models, including the 10-fold cross-validation approach **(**k = 10) for MLM-KHNN and Bayesian backpropagation training for ANN. Notably, the neural network was structured with 20 hidden layers and trained over 1000 epochs to ensure comprehensive learning. While this deep architecture enhances accuracy, it also necessitates significant computational resources, highlighting a key challenge for real-time applications.

Evaluation Metrics

To evaluate the performance of the proposed MLM-KHNN+ANN model, various statistical metrics were employed, including accuracy, precision, specificity, sensitivity, F1-score, and AUC-ROC. The formulas for calculating these

[DOI: 10.18502/jabs.v15i2.18083]

Table 2. Parameter Settings

Method	Parameter	Value
MLM-KHNN	Data partitioning method	k-fold
	K in k-fold	K = 10
ANN	Activation function	Sigmoid
	Cost function	MSE
	Epochs	1000
	Hidden layers	20
	Training algorithm	Bayesian Backpropagation

Abbreviations: MLM-KHNN: Multiple Local Mean Vector-based k-Nearest Neighbor, ANN: artificial neural network

Table 3. Performance Evaluation Metrics for Classification

Metric	Formula
Accuracy	(TN + TP) / (TN + TP + FN + FP)
Precision	TP / (TP + FP)
Sensitivity	TP / (TP + FN)
Specificity	TN / (TN + FP)
F1-score	2 (Precision Sensitivity) / (Precision + Sensitivity)

metrics are provided in Table 3. As indicated in Table 3, accuracy measures the overall correctness of the classification, while precision assesses the proportion of correctly identified positive cases. Sensitivity and specificity evaluate the model's ability to correctly identify diseased and non-diseased cases, respectively. The inclusion of the F1-score and AUC-ROC ensures a balanced assessment, particularly given concerns about class imbalance in medical datasets.

Comparison of Proposed Method with Other Methods

To validate the effectiveness of the proposed model, a comparative analysis was conducted against conventional classifiers such as SVM and ANN, as well as advanced techniques including Random Forest, Gradient Boosting, and deep learning-based classifiers. The performance on the test set is summarized in Table 4. According to Table 4, the proposed MLM-KHNN+ANN model achieved the highest accuracy (0.9610), outperforming SVM (0.9348), ANN (0.9422), Random Forest (0.9512), and Gradient Boosting (0.9580). The model's perfect sensitivity score (1.0000) indicates its strong ability to detect Parkinson's disease cases without missing any positive instances. However, its relatively lower precision (0.8330) suggests a higher false positive rate, which could be addressed in future work through improved feature selection and enhanced data balancing.

In addition to test set performance, Table 5 presents the comparative results when the models were trained and evaluated on the entire dataset (training + test data). As seen in Table 5, the proposed model achieved an impressive accuracy of 0.9900, surpassing other methods such as Gradient Boosting (0.9663) and Random Forest (0.9592). It also recorded the highest specificity (0.9860), indicating its superior ability to correctly classify non-diseased individuals. The AUC-ROC score (0.9855) further confirms the robustness of the proposed approach, positioning it as a reliable diagnostic tool.

The exceptionally high sensitivity (1.0000) observed in the proposed method, as shown in Table 4 and Table 5, raises concerns regarding potential class imbalance in the dataset. To investigate this, techniques such as stratified k-fold cross-validation and data resampling were applied. Despite these adjustments, the proposed method consistently maintained high

[DOI: 10.18502/jabs.v15i2.18083]

Table 4. Comparison of Methods for the Test Set

Method	Accuracy	Precision	Specificity	Sensitivity	F1-score	AUC-ROC
SVM	0.9348	0.9311	0.9241	0.9754	0.9527	0.9650
ANN	0.9422	0.9502	0.9420	0.9500	0.9501	0.9685
Random Forest	0.9512	0.9485	0.9551	0.9604	0.9544	0.9723
Gradient Boosting	0.9580	0.9602	0.9613	0.9671	0.9636	0.9750
Proposed method	0.9610	0.8330	0.9510	1.0000	0.9091	0.9800

Abbreviations: SVM: support vector machines, ANN: artificial neural network

Table 5. Comparison of Methods for the Entire Dataset (Training + Test)

Method	Accuracy	Precision	Specificity	Sensitivity	F1-score	AUC-ROC
SVM	0.9501	0.9499	0.9574	0.9699	0.9598	0.9700
ANN	0.9435	0.9551	0.9387	0.9463	0.9507	0.9692
Random Forest	0.9592	0.9600	0.9621	0.9675	0.9637	0.9785
Gradient Boosting	0.9663	0.9681	0.9704	0.9732	0.9706	0.9821
Proposed method	0.9900	0.9600	0.9860	1.0000	0.9796	0.9855

Abbreviations: SVM: support vector machines, ANN: artificial neural network

sensitivity without significantly compromising overall accuracy. This suggests that the model is highly effective in detecting Parkinson's disease cases but may benefit from further validation on diverse datasets to ensure generalizability.

Comparison with Previous Research

To further contextualize the findings, the results of this study were compared with previous research on Parkinson's disease diagnosis. Studies employing methods such as SVM, ANN, and deep learning have reported accuracy levels ranging between 92% and 97% (2–3, 15). The proposed method, achieving 99% accuracy, surpasses existing methods by leveraging the hybrid approach of MLM-KHNN and ANN, which enhances the classifier's ability to learn complex patterns while mitigating issues related to local minima in neural networks. Despite these improvements, it is important to note that the increased sensitivity suggests the need for further validation on larger and more diverse datasets to confirm generalizability. Additionally, integrating explainability techniques, such as SHAP values, could provide deeper insights into feature importance within the model. The proposed MLM-KHNN+ANN model consistently outperformed conventional classifiers and advanced machine learning techniques in Parkinson's disease diagnosis. The exceptionally high sensitivity indicates its effectiveness in identifying diseased cases; however, further validation is required to ensure robustness. The inclusion of the F1-score and AUC-ROC metrics provided a more comprehensive evaluation, demonstrating the model's strong discriminative ability. Future work should explore additional deep learning techniques and feature selection strategies to further enhance the model's performance.

Discussions

This study presents a novel hybrid approach that combines Multiple Local Mean Vector-based k-Nearest Neighbors (MLM-KHNN) with an Artificial Neural Network (ANN) trained using Bayesian backpropagation for the diagnosis of Parkinson's disease. The proposed method significantly improves classification accuracy while addressing common limitations of traditional classifiers, such as susceptibility to noise and parameter sensitivity. One of the key advantages of the proposed method is its exceptional accuracy and sensitivity. The MLM-KHNN+ANN model achieved an accuracy of 99%, sensitivity of 100%,

[DOI: 10.18502/jabs.v15i2.18083]

precision of 96%, and specificity of 98.6%. The perfect sensitivity score indicates the model's ability to detect Parkinson's disease cases without missing any positive instances—an essential feature for early diagnosis.

Additionally, the model demonstrates strong robustness in handling non-linear and noisy data. Traditional k-NN methods often struggle with determining optimal neighborhood sizes and managing non-linear distributions; however, MLM-KHNN effectively addresses these issues by incorporating multiple local mean vectors, leading to improved classification performance in complex medical datasets. Another advantage lies in the model's consistent performance across varying data configurations. Unlike standard ANN models, which can be highly sensitive to initial parameter settings, the Bayesian backpropagation training employed in this study enhances generalizability and reduces the risk of overfitting.

Furthermore, the MLM-KHNN approach improves feature representation and enhances classification capacity, enabling more accurate differentiation between healthy individuals and those with Parkinson's disease. This capability contributes to higher AUC-ROC scores compared to conventional classifiers.

The proposed model offers significant practical applications in clinical settings. Its high sensitivity makes it a valuable tool for early Parkinson's disease detection, thereby facilitating timely medical intervention. Moreover, clinicians can leverage the model's predictions to tailor treatment strategies—such as selecting appropriate medications or identifying candidates for advanced therapies like deep brain stimulation. Another promising application is its integration with wearable and imaging technologies. The model can be embedded in wearable sensor systems for continuous patient monitoring or adapted for medical imaging applications to improve Parkinson's disease detection accuracy. Despite its promising performance, the proposed approach is not without limitations. One major challenge is its computational complexity, as the hybrid structure of MLM-KHNN and ANN demands substantial computational resources. Deploying this model in real-time applications may require hardware acceleration techniques, such as the utilization of GPUs or TPUs. Another limitation involves parameter optimization. Although Bayesian backpropagation mitigates overfitting, fine-tuning the model's hyperparameters remains a complex task. Future work could explore advanced optimization techniques, such as genetic algorithms or Bayesian optimization, to further refine performance.

Additionally, the observed perfect sensitivity score raises the possibility of class imbalance in the dataset. While stratified cross-validation was employed to address this issue, further validation using larger and more diverse datasets is essential to confirm the model's robustness and generalizability.

A comparative analysis with existing techniques underscores the advantages of the proposed model. Compared to Support Vector Machines (SVMs), which are effective for binary classification but often struggle with large, high-dimensional datasets, the proposed model maintains high accuracy while reducing dependence on manual feature engineering. Traditional ANNs frequently suffer from convergence issues related to local minima and instability in parameter tuning. In contrast, the Bayesian backpropagation used in the proposed model enhances learning efficiency and generalization. When benchmarked against advanced classifiers such as Random Forest and Gradient Boosting, the proposed method outperformed both in terms of sensitivity and F1-score, demonstrating superior disease detection capabilities. Moreover, while deep learning approaches typically require large amounts of labeled data and substantial computational power, the proposed hybrid model delivers

[DOI: 10.18502/jabs.v15i2.18083]

comparable performance while remaining relatively computationally efficient.

Future directions for optimizing the model include incorporating feature selection algorithms to improve computational efficiency and boost classification accuracy. Advanced optimization methods, such as genetic algorithms or swarm intelligence-based optimizers, could be employed to fine-tune hyperparameters more effectively. Additionally, integrating hybrid deep learning architectures—such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs)—may further enhance the model's capacity to capture temporal or spatial patterns in medical datasets. Applying extensive k-fold cross-validation procedures would also help ensure the model's reliability across diverse population groups and data sources. Future research could explore adapting the model for the diagnosis of other neurodegenerative diseases, including Alzheimer's disease and multiple sclerosis. Moreover, integrating the model into AI-powered clinical decision support systems could enable real-time diagnostics in smart healthcare environments. Finally, incorporating genetic and lifestyle factors into the model could facilitate the development of personalized medicine applications, tailoring treatment plans to individual patient profiles.

Conclusion

The hybrid MLM-KHNN+ANN model presented in this study offers a highly effective approach to Parkinson's disease diagnosis, achieving superior classification accuracy and robustness compared to both traditional and advanced classifiers. The model's exceptional sensitivity and specificity underscore its value as a tool for early disease detection. Future enhancements—such as feature selection, parameter optimization, and the integration of deep learning techniques—could further refine its performance and broaden its applicability in real-world clinical environments. With additional validation using larger and more diverse datasets, this model holds the potential to significantly impact medical diagnostics and personalized healthcare solutions.

Acknowledgment

The author would like to express sincere gratitude to the Islamic Azad University, Estehban Branch, for their cooperation and support throughout the research process. This study has been officially registered at the Islamic Azad University, Estehban Branch, under the registration code 1457296247162616560141627295 00.

Conflict of Interests

No conflicts of interest have been declared by the author.

Funding

This research did not receive any financial support from public, commercial, or non-profit funding agencies.

Ethical Considerations

This study utilized publicly available data from the UCI Machine Learning Repository, ensuring full compliance with ethical standards and data privacy regulations. No human subjects were directly involved, and the research adheres to ethical guidelines for the use of secondary data. The dataset link is provided in the Data Availability Statement section. Any use or reproduction of this research is permitted with the author's prior consent.

Authors' Contributions

All the research implementation and analysis of the results and writing of the article were done by Mohammad Javad Hosseinpour.

Data Availability Statement

The dataset is available on the UCI website according to the link below: https://archive.ics.uci.edu/dataset/174/parkinsons

[DOI: 10.18502/jabs.v15i2.18083]

References

1 Rezaee K, Zhu M. Diagnose Alzheimer's disease and mild cognitive impairment using deep CascadeNet and handcrafted features from EEG signals. Biomed Sign Proc Control. 2025; 99(1): 106895.
2 Islam M A, Majumder M Z H, Hussein M A, Hossain K M, Miah M S. A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets. Heliyon. 2024; 10(3): 25469.
3 Ahmadi H, Huo L, Arji G, Sheikhtaheri A, Zhou S M. Early diagnosis of Parkinson's disease using a hybrid method of least squares support vector regression and fuzzy clustering. Biocybern Biomed Eng. 2024; 44(3): 569-585.
4 Jiang M, Liu Y, Cao Y, Liu Y, Wang J, Li P, et al. Auxiliary diagnostic method of Parkinson's disease based on eye movement analysis in a virtual reality environment. Neurosci Lett. 2024; 842(1): 137956.
5 Desai S, Chhinkaniwala H, Shah S, Gajjar P. Enhancing Parkinson's disease diagnosis through deep learning-based classification of 3D MRI images. Procedia Comput Sci. 2024; 235(1): 201-213.
6 Hosseinpoor M. Intelligent diagnosis of heart diseases based on electrocardiograph signal. *J Adv Biomed Sci*. 2024; 14(1): 47-55.
7 Khosla A, Kumar N, Khera P. Machine learning approach for predicting state transitions via shank acceleration data during freezing of gait in Parkinson's disease. Biomed Sign Proc Control. 2024; 92(1): 106053.
8 Esfahani M M, Mostashfi M, Hosseinabadi S V, Hashemi M S, Peymani M, et al. Unveiling the regulatory role of miR-101-3p on ZNF746 in a Parkinson's disease cell model: Implications for therapeutic targeting. Neurosci Res. 2024; 203(1): 18-27.
9 Hosseinpoor M. Presenting an intelligent algorithm to predict heart attack. J Health Biomed Inform. 2024; 10(4): 346-356.
10 Pasha A, Ahmed S T, Painam R K, Mathivanan S K, Karthikeyan P, Mallik S, et al. Leveraging ANFIS with Adam and PSO optimizers for Parkinson's disease. Heliyon. 2024; 10(9): 30241.
11 Xu M, Jin H, Ge W, Zhao L, Liu Z, Guo Z, et al. Mass spectrometric analysis of urinary Nglycosylation changes in patients with Parkinson's disease. ACS Chem Neurosci. 2023; 14(18): 3507-3517.
12 Hasanzadeh M, Mahmoodian H. A novel hybrid method for feature selection based on gender analysis for early Parkinson's disease diagnosis using speech analysis. Appl Acoust. 2023; 211(1): 109561.
13 Abumalloh R A, Nilashi M, Samad S, Ahmadi H, Alghamdi A, Alrizq M, et al. Parkinson's disease diagnosis using deep learning: A bibliometric analysis and literature review. Ageing Res Rev. 2024; 96(1): 102285.
14 Sigcha L, Borzì L, Amato F, Rechichi I, Romero C R, Cárdenas A, et al. Deep learning and wearable sensors for the diagnosis and monitoring of Parkinson's disease: A systematic review. Expert Syst Appl. 2023; 229(1): 120541.
15 Tanveer M, Rashid A H, Kumar R, Balasubramanian R J I P. Parkinson's disease diagnosis using neural networks: Survey and comprehensive evaluation. Inf Process Manag. 2022; 59(3): 102909.
16 Vyas T, Yadav R, Solanki C, Darji R, Desai S. Deep learning-based scheme to diagnose Parkinson's disease. Expert Syst. 2021; 39(3): 12739.
17 Rajalaxmi R R, Mirjalili S, Gothai E, Natesan P. Binary grey wolf optimizer with mutation and adaptive k-nearest neighbour for feature selection in Parkinson's disease diagnosis. Knowl-Based Syst. 2022; 246(1): 108701.

[DOI: 10.18502/jabs.v15i2.18083]